PROBLEMS - SEVEN LECTURES IN ALGEBRAIC STATISTICS

ELIANA DUARTE

Some of these problems require you to use either the statistical software `R` or a symbolic algebra software. My recommendation for algebra software is `Macaulay2`, but feel free to use any other that works for you.

We will carry out Fisher's exact test in `R`. To be able to this computation you need to install several packages including the package `algstat`. Go to https://github.com/dkahle/algstat for instructions on how to get these packages.

Exercise 1 (Ex. 6.1 [1]). Recall that for $\lambda > 0$, a random variable X has a Poisson distribution with parameter λ if its state space is $\mathbb{N}$ and

$$P_\lambda(X = k) = \frac{\lambda^k}{k!} \exp -\lambda \text{ for } k \in \mathbb{N}.$$

Show that the family of Poisson random variables is a regular exponential family. What is the natural parameter space?

Exercise 2 (Ex 6.2 [1]). Consider the vector $h = (1, 1, 1, 2, 2, 2)$ and the matrix

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 & 1 & 1 \end{pmatrix}$$

1. Compute generators for the toric ideals I_A and $I_{A,h}$.
2. What familiar statistical model is the discrete exponential family $\mathcal{M}_{A,h}$.
3. Given a vector $u \in \mathbb{N}^6$ what is the interpretation of the vector of sufficient statistics Au?

Exercise 3. Consider the table u below that classifies 326 homicide indictments in Florida under two binary criteria: $X =$race and $Y =$death penalty.

	Death P	No Death P	Total
Black	19	141	160
White	17	149	166
Total	36	290	326

Consider the question: Where death penalty decisions made independent of race? This is a hypothesis test: $p \in \mathcal{M}_{X \perp\!\!\!\perp Y}$ vs $p \notin \mathcal{M}_{X \perp\!\!\!\perp Y}$. Recall that the matrix A for the independence model is

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

1. Compute a basis for the toric ideal of the model $\mathcal{M}_{X \perp\!\!\!\perp Y} = \mathcal{M}_A$. Write the elements of the basis as 2×2 contingency tables.
2. Use R to do Fisher's asymptotic test and Fisher's exact test.
3. How many tables are in the fiber $\mathcal{F}(u)$?
4. What happens when you change the values in the entries of the table to values that are smaller than 3?

Exercise 4 (Ex. 4.7 [1]). For four binary random variables, consider the conditional independence model $\mathcal{C} = \{1 \perp\!\!\!\perp 3 \mid \{2,4\}, 2 \perp\!\!\!\perp 4 | \{1,3\}\}$. Compute the primary decomposition of $I_{\mathcal{C}}$ and describe the components.

Exercise 5 (Ex. 4.10 [1]). Consider the marginal independence model for four binary random variables $\mathcal{C} = \{1 \perp\!\!\!\perp 2, 2 \perp\!\!\!\perp 3, 3 \perp\!\!\!\perp 4, 1 \perp\!\!\!\perp 4\}$.

1. Compute the primary decomposition of $I_{\mathcal{C}}$
2. Show that there exists exactly one component of $V(I_{\mathcal{C}})$ that intersects the probability simplex.
3. Give a parameterization of that component.

References

1. S. Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.

Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany
E-mail address: eliana.duarte@ovgu.de